DOCUMENT RESUME

ED 214 134 CS 006 553

AUTHOR Athey, Irene
TITLE Reading Research Synthesis: Problems and Challenges.
PUB DATE Mar 82
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Graduate Study; Higher Education; *Literature Reviews; Reading Instruction; *Reading Research; *Research Methodology; *Synthesis

ABSTRACT

Among the advantages of a well-written review of the reading research literature are that it lays down the framework of the topic for the reader and provides exercise for both the synthesizer and the reader in integrating widely scattered, incompatible research findings. These benefits are intimately related to the problems in such syntheses, such as the limitations and loopholes of a computer data base search and the fallibility of the researcher's judgment, selection, and classification systems. Much writing and rewriting of the research review are required to ensure clarity, sustained interest, and balance. Most of what takes place in developing a research synthesis is descriptive, rather than prescriptive. Research synthesis, however, should play a prominent role in the prescriptive domain in order to provide both the descriptions of reading behavior and the design for changing that behavior through educational practice. Recommendations based on such syntheses should generate more confidence in their scientific underpinnings than do single studies. Such reviews appear infrequently in the reading research literature because few students receive systematic training in conducting research syntheses and because such work tends to be undervalued in comparison with "pure" research. (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED214134

Reading Research Synthesis: Problems and Challenges

Irene Athey
Rutgers University

Reviewing the literature is an exercise that every graduate student must undertake in the course of writing a doctoral dissertation or a master's thesis. Many spend as much as three to six months on this phase, but it seems that few receive specific instruction in the art of collecting and synthesizing research material. Perhaps we all assume that students get plenty of practice in reviewing the literature by writing term papers in their courses, or simply that an intelligent being who knows how to use the subject index in the library can proceed from there and with the requisite amount of work come up with a polished manuscript that says it all in 25 pages.

Experimental research is based on a scientific model, though it is also to some extent an art. Similarly, research synthesis is both a science and an art, though possibly a different kind of art from experimental research itself. Certainly there are points of overlap, both in methods and problems. Before discussing these, I would like to expound for a minute or two on the virtues of research syntheses, assuming, of course, that these syntheses meet the criteria which we shall presently lay down for them. I do this because I have a sense that, within the research establishment, research synthesis occupies a second-class status, and that those who do it (especially if they do it exclusively) are

---

Invited address for Division C of the American Educational Research Association, New York, March 20, 1982.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Irene Athey
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS006353

regarded as engaging in a parasitic activity, living off the brains and work of others and lacking the creativity or the will power to do the real thing. I hope that my remarks, if they accomplish nothing more, will dispel the illusion that research synthesis is an easy or a lame-brain activity.

The major advantage of a well-written review is that it lays down the parameters, or provides the framework, of the topic for the reader. This may sound obvious, but its importance cannot be overemphasized. In recent years, we have come to understand that, in reading a passage or a story, a reader will elaborate or distort information in order to make it meaningful in terms of what is already known about the subject, and that it is therefore incumbent on the author to provide some kind of initial framework to the discussion in order to guide the assimilation and organization of the new information on the part of the reader. A good research synthesis serves the same function in introducing the reader to a whole area of knowledge in the same way that the opening sentences of a passage assist the reader to draw the boundaries of discourse to be expected.

Clearly, understanding a whole research area is a long-drawn out process that is usually inductive in nature, that is to say, the typical student comes to acquire this understanding in piecemeal and additive fashion, reading first this study, later another one, attempting to reconcile them perhaps, until something more definitive comes along. The conceptual framework needed to make sense of this piecemeal material is usually developed after much energy has been expended in fitting the

pieces together. But once a framework begins to develop, from there on out, the process becomes deductive, a matter of fitting each new piece of information into the organized structure of knowledge about the topic. A research synthesis provides such a framework, and hence reduces the period of inductive research while increasing the meaningful assimilation of each new item as it appears.

The second advantage of research synthesis is that it presents a balanced treatment of the issues that have plagued, or at least enlivened debate and stimulated research in the area. The perspective that this balanced treatment brings to the subject can range from understanding of a particular argument between two researchers to an appreciation of how the specific issue fits into the historical perspective, the overarching philosophical argument that has been going on for centuries. Hence, a good review should clarify the origins of the dispute, stages in its development, how specific research studies have advanced our thinking about the issue and, last but not least, the status of the issue at the time of writing. Such forays into the past often reveal some interesting historical information. For example, I remember doing an experiment once in a lab course at the University of California in which I used the Stroop Color Test. I decided to investigate the origins of this test, which was not very well-known at the time, and was surprised to discover that considerable controversy raged among psychologists during the 30s over the meaning of this test. I learned, among other things, that issues that seem to be of great moment and are very divisive at the time, tend to dissipate and become forgotten with the passage of time. Still, while they exist, they exist for a good reason, and the resolution of the conflict and its disappearance from the literature may represent

a giant step forward for psychological science.

Where annual reviews are a feature of the discipline, as is the case with the annual summaries of investigations relating to reading that, until recently appeared for many years in the Reading Research Quarterly they can serve as a historical record of the directions this discipline has taken. A simple perusal of the title headings can serve to show the rise of one area of interest and the decline of another, or, in some cases, continued neglect of an area. The role of affective factors in reading would be a case in point. To illustrate this point, about two years ago, I made a tally of all abstracts appearing in the annual summaries for three successive years (1977-79), and found that, using the most liberal criterion possible, out of a total total of 1,791 references, only 144 were related to what I consider to be an important domain in the analysis of reading processes, not to mention the teaching of reading.

Third, it seems to me, the research synthesis is valuable in broadening the reader's concept of suitable methodologies for addressing the problem under consideration. In pursuing a line of research, it is very easy to become wedded to a particular mode of research design or statistical technique. We justify this action by saying that it is easier to compare studies that employ the same method, and hence more possible to draw valid conclusions from the accumulated evidence, and this is true. On the other hand, we all bring our personal epistemologies, philosophical and scientific beliefs to our work, with all

the implicit biases entailed therein. A survey of research, especially if it involves a foray into another discipline, can introduce a whole new way of looking at the problem, though this too has its own dangers as we shall see presently.

Perhaps the greatest value of all to be gained from a well conducted research synthesis is the exercise it provides for both the synthesizer and the reader in integrating widely scattered, incompatible, sometimes incomprehensible research findings. The ability to accomplish such an integration is an art not to be underestimated, in my opinion. But when it is accomplished, the reader can be led through the same sequence of problem-solving steps that were followed by the author. I view this as a valuable intellectual exercise over and above the content that is transmitted.

My assignment today however, was to address not the advantages of research synthesis, but its problems and challenges. I felt it would be a little unfair to dwell on the problems without mentioning the benefits, and in any case, my discussion of the advantages leads rather nicely into the problems, because they are, as you will see, intimately related. Let me return to the first point, the establishment of a framework that a good synthesis provides for the reader. This framework must be constructed, and is not achieved without considerable thought, or even creative imagination, on the part of the synthesizer. Picture yourself coming to this task of making sense of an unknown number of articles scattered throughout the journals. Of course, the first task is to find out exactly what is out there, and most people

go about that by initiating a computer search. Mechanical as this may seem, a conceptual process is brought into play by the very choice of key words. If the key words define the domain of discourse too narrowly, much relevant material will not surface. If they define it too broadly, much of it will be unusable. The latter possibility is increased if one is investigating the conjunction of two fields or areas of research, each with its own extensive literature. At this point, it is better to err on the side of being too broad. The net must be cast widely, to avoid the error of exclusion. By the way, one can never be completely confident that one's search has been exhaustive, however wide the net. I once exchanged computer-generated lists with a fellow researcher who had tapped into a different system, and discovered only about a 30 percent overlap in our two lists. Obviously, what comes out of the hopper cannot be greater than what is in there to start with. So it is necessary to supplement the computer with one's own checks on the indices of leading journals and other sources of information.

Now comes the inductive part of the investigation. It may not be possible to read everything that the computer has churned out, but the conscientious reviewer must consider the merits of every article, either by skim reading, reading the abstract, or assigning someone else to read the article and make a careful abstract. A preliminary decision to include or exclude may be made at this point simply on the grounds of intrinsic merit, without as yet any firm guiding principles. As one continues to read, these principles or guiding concepts will begin to emerge, so that a tentative classification of the material can be made.

It is these concepts that will determine the direction of the search thereafter. Generally speaking, as I have indicated, these concepts are arrived at inductively. The nature of the research material "out there" determines to some extent how one categorizes it. Sometimes it is possible, however, to entertain some hypothesis, and to organize the material so that it sheds some light on the merits of the hypothesis. For example, in my disseration, I was playing with the idea that some of Erik Erikson's concepts as to what constitutes a healthy personality, concepts such as mastery of the environment, freedom from excessive anxiety, etc., might be related to reading achieveement. In reviewing tne literature on reading in relation to the affective domain, I used these concepts to categorize the studies, being careful, of course, to include studies that both did and did not support the hypothesis. The problem with this approach is that not all studies will fit neatly into the classification system, and judgments must be made as to whether excluding those that don't is likely to bias the outcome. Another example of guiding principles being given a priori is the Targeted Research conducted for the U.S. Office of Education in the early 70s. In this case, the thinking was that too many unrelated empirical studies on factors related to reading achievement had been conducted without any consistent outcome or conclusion that merited a degree of confidence. The initiators of this enterprise felt strongly that the time had come to test existing models, or even partial models, of the reading process by ascertaining whether and to what degree studies already conducted supported these models. Again, this approach provided an interesting intellectual exercise, first in identifying

the models (a problem in itself), and second in determining at what points the research literature matched particular features of the models.

We will assume that our researcher has been intellectually honest in deciding, on the basis of some principles or concepts, what to include in the review, and now has a corpus of literature on which to work. Now the problems really begin. To start with the studies will not distribute themselves neatly and evenly across the categories. There will be holes in the data, often significant holes. It is not like going out and gathering one's own data, where one can (more or less) determine what data will be collected, and why. The task is more like using someone else's data bank. If you have ever tried to do this, you will know that one usually finds that those data were collected for a different purpose, and that the data one really needs are missing. By the same token, the reviewer cannot manufacture studies that have not been conducted however glaring the omission may seem. Needless to say, it is very discouraging to find such lacunae, and may, if serious enough, result in discarding one's classification system in favor of a new one that more readily accommodates the material in hand.

Let us suppose that the reviewer is now in possession of a manageable body of research literature, has a reasonably satisfactory classification scheme, and is ready to start the process of reviewing and integrating it. Every study must be carefully examined and its inherent weaknesses and limitations noted. I speak here not only of such obvious factors as the inappropriate use of statistical techniques, or the overgeneralization of conclusions, and all the other criticisms one regularly

hears in dissertation oral exams, but also of such considerations as whether the study is indeed a test of the theory it claims to be testing. This is a very complicated matter; a conclusion that a study is not a valid test may be warranted for any one of a number of reasons. First, the theory may be essentially untestable, because the key concepts are vague or untranslatable into empirical terminology. In other cases, the concepts may be fully explained in the theory, but the measurement instrument devised to operationalize them may deviate substantially from the original intent of the theorist. (I have never been fully satisfied, for example, that anxiety, as measured by the Taylor Manifest Anxiety Scale, was indeed a valid measure of drive, as conceived in the Hullian theory, although it was so used in a number of experiments.) Finally, the hypothesis supposedly derived from the theory may not logically derive from it all. Indeed, sometimes the logic appears to be quite tenuous. The reviewer must be critically aware of all these possibilities, especially when the conceptual framework involves accumulating evidence to test a theory or model, as was the case with the Targeted Research. Obviously, when a researcher sets up a study to test a hypothesis derived from a theory, some conclusion is going to be drawn as to whether the theory stands up to the test or needs to be modified in some way. These conclusions may be quite fallacious, if the logic described in the above steps is not watertight.

By this time, the reviewer has been working on the process for several weeks, or even months. It is necessary to review the recent journals periodically to be sure that any relevant new studies are

included. (One hopes fervently that a synthesis of the literature does not appear during this period.) The period may be quite extended, because the work tends to be tedious, and because much thinking about the studies and how they fit into the conceptual framework is required. But eventually it becomes time to start integrating the studies in order to generate some worthwhile conclusions. This phase can be particularly exasperating. It is not enough to note the weaknesses of the individual studies; one must assess the relative damage they inflict on the generalized conclusions. But let us suppose by some happy quirk of fate that the studies are all well-conceived and carried out. It may still be impossible to integrate them because of the fact that different populations, different instruments, and different methodologies were used. These differences become problematic when the conclusions are in conflict. Are the contradictory findings due to the fact that different measures were used to operationalize self-concept, or because the samples differed in age or ethnic background, or because the treatments differed in length and intensity? If by some fortunate chance these are all the same, and the conclusions still differ, one must look for the reasons elsewhere in the details of each study. Obviously, the more studies there are to be integrated, the more complex the task, unless by some happy circumstance, the studies are all of high quality and the trends are largely clear and in the same direction.

Writing a research review poses its own problems. In articles where a single study is being reported, the research review is usually terse and non-explanatory, because the purpose is to provide just enough

information to demonstrate the background from which the study evolved. By contrast, the research synthesis must draw a fine balance in the amount of information provided for any one study. Apart from considerations of fairness and objectivity, the author must give just enough information to outline the dimensions of the study without becoming so involved in detail that the thread of the argument is lost, or the article itself become inordinately long. Much writing and rewriting is required at this stage to ensure clarity of exposition, sustained interest on the part of the reader, and a balanced treament of the research cited.

I come now to a very important function of research synthesis. To explicate it I will call on the distinction made by Robert Glaser (1976) between "descriptive" and "prescriptive" science. Although conceptual and judgmental activities are assuredly involved in the processes we have described, by and large the bulk of what takes place in developing a research synthesis falls under the "descriptive" category. We often speak of the need to "bridge the gap" between theory and practice, and Glaser reiterates that need by calling for a "linking science" -- the psychology of instruction -- that is "prescriptive", and in which the primary intellectual activity is that of design. In his book, The Sciences of the Artificial (1969), Simon points out that it is the task of the professions to prescribe ways of designing products such as buildings or curricula. In fact, Glaser maintains, "there is, at the present time, a growing feeling that a strong test of the adequacy of descriptive theory in the behavioral and social sciences can be made through attempts at application based upon the development of

prescriptive theory for the design of social policy and social instiitutions, including education" (p.5).

I would like to argue that syntheses of research could and should play a prominent role in this prescriptive science, that they are in a strong position to provide both the descriptions of behavior and the design for changing that behavior through educational practice. I recognize that most authors of single research studies usually attempt to extrapolate their findings to practice, but I seriously wonder if such extrapolation is warranted. On the other hand, research syntheses provides much stronger justification for making recommendations and pointing out directions that instructional practices might take. Recommendations based on a synthesis of the research should presumably generate more confidence in their scientific underpinnings than single studies alone. This being the case, it is incumbent on the writer of syntheses to offer some suggestions for instruction based on the weight of corroborative evidence incorporated in the review. A conceptually rigorous and cogent review can thus have the dual effect of changing the direction of research (descriptive) and instructional practice (prescriptive) at the same time. This two-fold impact emphasizes the heavy responsibility of the synthesizer to draw conclusions with extreme care, taking full cognizance of all the compounding factors we have discussed. The apparent weight of the accumulated evidence should not deceive the writer into propounding the conclusions with unwarranted confidence and vigor.

I had planned to include among the "problems and challenges" a paragraph about the timing of research syntheses. What is a good time to embark on such a venture? The answer to that question presupposes a certain degree of familiarity with the field on the part of the reviewer and a specific purpose for undertaking such an arduous and time-consuming task. When a topic is "hot", as we say, the time may seem opportune for such a review; on the other hand, it may be very difficult to ensure the necessary distance and degree of objectivity, especially if one is caught up in conducting research on that topic. It may seem advisable under these circumstances to let the dust settle a little before trying to shed light on what is happening. On the other hand, if one is concerned about the direction research is taking, and believes it to be wrong-headed or fruitless, then a timely research review may have the desired effect of changing its course. Finally, a good synthesis may be instrumental in drawing attention to an area that has been much neglected, but that the author deems important. So the timing of a research synthesis would appear to be quite important.

On further reflection, however, I decided not to highlight that part of the discussion, lest some aspiring reviewer be discouraged from starting such a review. The truth is that we simply do not have enough syntheses of the literatre on any given topic at any time. We need to be constantly apprized of the state of the art. We need to know what all those research studies being conducted really amount to, and how their combined evidence advances the state of

descriptive science. Not only is evidence constantly accumulating, but this very accumulation changes the conceptual framework we impose on that evidence. If I were to conduct another review on language factors related to reading, for example, it would be very different from the one that appeared ten years ago. The conceptual framework has changed dramatically.

Last but not least, our information needs to be continuously updated on what the integration of this knowledge allows us to prescribe for the psychology of instruction. Thus, an excellent review that incorporates all these features is never untimely.

Why are there not more reviews appearing regularly in the literature? I think, for a number of reasons. As I noted in my opening remarks, few students receive systematic training in conducting research syntheses. Moreover, their brief exposure to the process through term papers and dissertation reviews gives them an inkling of the amount of work and the problems involved. But mostly because such work tends to be undervalued relative to the conduct of "pure" research. It is true that students are assigned to read and may receive inspiration from such "classic" reviews as Robert White's "Motivation reconsidered" that have had a significant impact on the psychological thinking of the period, but it may be difficult for a student reader to identify with such models. Yet many, I feel, would willingly testify to the value they have received from such articles. Unless we ourselves are convinced of the value of this intellectual exercise, it seems unlikely that the next generation of scholars will be any more prolific in this regard than we are ourselves, to the detriment of both the descriptive and the prescriptive science.

References

Athey, I., Reading: The affective domain revisted. Paper delivered at C. W. Post Center, Long Island University, August 1980.

Athey, I., Language models and reading. Reading Research Quarterly, 1971, 7(1), 16-110.

Glaser, R., Components of a psychology of instruction: Toward a science of design. Review of Educational Research, 1976, 46(1), 1-24.

Simon, H. A. The Sciences of the Artificial. Cambridge, Mass.: M.I.T. Press, 1969.

White, R. W., Motivation reconsidered: The concept of competence. Psychological Review, 1959.